Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 493–505
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Optimized variable source-profile approach for source apportionment

Amit Marmur*, James A. Mulholland, Armistead G. Russell

School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0512, USA

Received 9 May 2006; received in revised form 10 August 2006; accepted 19 August 2006

Abstract

An expanded chemical mass balance (CMB) approach for $PM_{2.5}$ source apportionment is presented in which both the local source compositions and corresponding contributions are determined from ambient measurements and initial estimates of source compositions using a global-optimization mechanism. Such an approach can serve as an alternative to using predetermined (measured) source profiles, as traditionally used in CMB applications, which are not always representative of the region and/or time period of interest. Constraints based on ranges of typical source profiles are used to ensure that the compositions identified are representative of sources and are less ambiguous than the factors/sources identified by typical factor analysis (FA) techniques. Gas-phase data (SO_2, CO and NO_y) are also used, as these data can assist in identifying sources. Impacts of identified sources are then quantified by minimizing the weighted-error between apportioned and measured levels of the fitting species. This technique was applied to a dataset of $PM_{2.5}$ measurements at the former Atlanta Supersite (Jefferson Street site), to apportion $PM_{2.5}$ mass into nine source categories. Good agreement is found when these source impacts are compared with those derived based on measured source profiles as well as those derived using a current FA technique, Positive Matrix Factorization. The proposed method can be used to assess the representativeness of measured source-profiles and to help identify those profiles that may be in significant error, as well as to quantify uncertainties in source-impact estimates, due in part to uncertainties in source compositions.

Keywords: CMB-LGO; Optimization; Source-apportionment; $PM_{2.5}$; PMF; Health-study

1. Background

Chemical mass balance (CMB) receptor models are a common tool for apportioning ambient levels of pollutants (mainly particulate matter) among the major contributing sources. CMB combines the chemical and physical characteristics of particles or gases measured at sources and receptors to quantify the source contributions to the receptor. Quantification is based on the solution to a set of linear equations that express each receptor's ambient chemical concentration as a linear sum of products of source-profile fractions and source contributions (US-EPA, 2004a, b), as expressed by

$$C_i = \sum_{j=1}^{n} f_{ij} S_j + e_i, \quad (1)$$

where C_i is the ambient concentration of chemical species i ($\mu g\,m^{-3}$), $f_{i,j}$ the fraction of species i in

*Corresponding author. Tel.: +1 404 385 4565; fax: +1 404 894 8266.
E-mail address: amit.marmur@ce.gatech.edu (A. Marmur).

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2006.08.028

emissions from source j, S_j the contribution (source-strength) of source j ($\mu g\,m^{-3}$), n the total number of sources, e_i the error term.

Source profile fractions (f_{ij}) and the receptor concentrations (C_i), along with uncertainty estimates, serve as input data to the CMB model. Results consist of the contribution of each source category (S_j) to the measured concentration of different species at the receptor. A frequent source of uncertainty in the implementation of CMB is the choice of source profiles used as input. There is a wide variety of source profiles in the literature, but these are not always representative of the region and/or time of interest. Some examples of this are as follows: soil (dust) composition often varies geographically; emission composition from biomass burning is dependent on the type of vegetation or wood burned (e.g. agriculture burning, soft or hard wood residential combustion); emissions from coal-fired power plants may vary depending on the types of coal used; mobile source emissions can vary from region to region and temporally due to different fuels, fleet composition, or driving conditions. To date, the most common approach to addressing this variability has been to select profiles that are most representative of the region and time period of interest from those that are available. In many cases, however, specific profiles are not available. Moreover, a profile derived from any one source at one time may not be representative due to variability in time and space. Due to these reasons, factor analysis (FA) techniques have been developed (Hopke, 1988; Paatero and Tapper, 1994) and are often applied to characterize and quantify the sources contributing to ambient particulate matter levels (Kim et al., 2003; Kim et al., 2004a, b; Maykut et al., 2003). FA models do not require the use of predetermined source profiles, but results are often difficult to interpret as factors do not necessarily represent specific sources (Seinfeld and Pandis, 1998). The underlying assumption in all FA models is that the chemical composition of ambient particulate samples includes information about the fingerprints of the sources affecting the receptor, and that this information can be used to derive the source compositions. The procedure for characterizing these sources (or factors) is based on correlations between ambient levels of the different species, a high correlation indicating that the species share a common source (Seinfeld and Pandis, 1998). One of the more commonly used FA methods in recent years is Positive Matrix Factorization (PMF) (Paatero and Tapper, 1994). In PMF, factors are constrained to have non-negative $f_{i,j}$'s, and no sample can have a negative source contribution. Application of PMF requires that error estimates for the data be chosen judiciously so that the estimates reflect the quality and reliability of each of the data points. A critical step in PMF analysis is the determination of the number of factors (Paatero, 2004).

2. Methods

2.1. CMB model expansion to include variable source compositions

This study combines concepts from FA and CMB applications to calculate source contributions to ambient $PM_{2.5}$ without relying solely on emissions composition studies or on interpretation of factors obtained by FA as sources. The technique is based on solving the same set of equations used in CMB modeling Eq. (1), but instead of using predetermined source profiles, ranges for different fractions in source-indicative profiles are used as input. The model then optimizes the fractions of different species within each profile by minimizing residual mass, subject to several constraints. Lower and upper bounds for the fractions of species in the various source profiles are set based on knowledge of typical compositions of various sources. Instead of deriving the contributing factors by FA, and then identifying (interpreting) them as sources based on knowledge of typical composition, this information is used beforehand to constrain the model while searching for the best combination of sources to describe the ambient levels of $PM_{2.5}$. The choice of source categories to include likewise is made beforehand, in contrast to FA.

As a basis for setting the constraints for the fractions of various species in the source profiles, suggested values are taken from in the validation protocol for CMB8.2 (US-EPA, 2004b) (Table 1). These bounds on the abundance of species were slightly modified (Table 2) and several additional constraints were added to better characterize the different sources. Emissions from light-duty gasoline vehicles (LDGVs) usually contain more OC than EC (Gillies and Gertler, 2000), so a constraint of $OC/EC \geqslant 1$ was used for LDGVs, and an opposite constraint was used for heavy duty diesel vehicles (HDDVs). However, the relative amount of EC and OC components in PM emissions from both

Table 1
Chemicals from particles in different emissions sources (US-EPA, 2004b)

Source type	Dominant particle size	Chemical fractions			
		<0.001	0.001–0.01	0.01–0.1	>0.1
Motor vehicles	Fine	Cr, Ni, Y	NH_4^+, Si, Cl, Al, Si, P, Ca, Mn, Fe, Zn, Br, Pb	Cl^-, NO_3^-, SO_4^{-2}, NH_4^+, S	OC, EC
Vegetative burning	Fine	Ca, Mn, Fe, Zn, Br, Rb, Pb	NO_3^-, SO_4^{-2}, NH_4^+, Na^+, S	Cl^-,K^+, Cl, K	OC, EC
Coal-fired boiler	Fine	Cl, Cr, Mn, Ga, As, Se, Br, Rb, Zr	NH_4^+, P, K, Ti, V, Ni, Zn, Sr, Ba, Pb	SO_4^{-2}, OC, EC, Al, S, Ca, Fe	Si
Soil dust	Coarse	NO_3^-, NH_4^+, P, Zn, Sr, Ba	SO_4^{-2}, Na^+, K^+, P, S, Cl, Mn, Ba, Ti	EC, OC, Al, K, Ca, Fe	OC, Si

Table 2
Lower and upper bounds for chemical fractions of total $PM_{2.5}$ mass emitted in source profiles

Species	Gasoline vehicles (GV)		Diesel vehicles (DV)		Soil dust (DUST)		Vegetative burning (BURN)		Coal power plants (CFPP)	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
EC	0.05	1	0.4	1	0	0.01	0.01	0.3	0.01	0.1
OC	0.3	1	0.1	1	0.01	0.1	0.3	1	0.01	0.3
SO_4^{-2}	0.01	0.1	0.01	0.1	10^{-3}	0.01	10^{-3}	0.01	0.01	0.3
NO_3^-	0.01	0.1	0.01	0.1	0	10^{-3}	10^{-3}	0.01	0	0
Cl^-	0.01	0.1	0.01	0.1	10^{-3}	0.01	0.01	0.1	0	0.01
NH_4^+	10^{-3}	0.01	10^{-3}	0.01	0	10^{-3}	10^{-3}	0.01	10^{-3}	0.01
Al	10^{-3}	0.05	10^{-4}	10^{-3}	0.01	0.2	0	0	0.01	0.1
As	0	0	0	0	0	0	0	0	0	10^{-3}
Ba	0	0	0	0	10^{-3}	0.01	0	0	10^{-3}	0.01
Br	10^{-4}	0.01	10^{-5}	10^{-3}	0	0	0	10^{-3}	0	10^{-3}
Ca	10^{-3}	0.05	10^{-4}	10^{-3}	0.01	0.2	0	10^{-3}	0.01	0.2
Fe	10^{-3}	0.05	10^{-4}	10^{-3}	0.01	0.2	0	10^{-3}	0.01	0.1
K	0	10^{-3}	0	10^{-4}	0.01	0.1	0.01	0.1	10^{-3}	0.01
Mn	10^{-4}	0.01	10^{-4}	10^{-3}	10^{-3}	0.05	0	10^{-3}	0	0.005
Pb	0	10^{-3}	0	10^{-3}	0	0	0	10^{-3}	0	0.01
Se	0	0	0	0	0	0	0	0	10^{-4}	0.01
Si	10^{-3}	0.05	10^{-4}	0.01	0.1	1	0	0	0.05	0.2
Ti	0	0	0	0	10^{-3}	0.05	0	0	10^{-3}	0.01
Zn	10^{-3}	0.02	10^{-4}	10^{-3}	0	10^{-3}	0	10^{-3}	10^{-3}	0.01
Additional constraints	Sum$\leqslant$1; OC/EC$\geqslant$1; TC$\geqslant$0.5; OM/OC$\geqslant$1.4		Sum$\leqslant$1; OC/EC$\leqslant$1; TC$\geqslant$0.5; OM/OC$\geqslant$1.4		Sum$\leqslant$1; Sum metal oxides$\leqslant$1		Sum$\leqslant$1; OC/EC$\geqslant$3; TC$\geqslant$0.5; OM/OC$\geqslant$1.4		Sum$\leqslant$1; Sum metal oxides$\leqslant$1;	

gasoline and diesel vehicles is highly variable (Gillies and Gertler, 2000), and there is significant overlap in the range of values between the two mobile source types. Therefore, trying to distinguish gasoline and diesel contributions separately on the basis of just EC and OC mass fractions is suspect (Gillies and Gertler, 2000). For this reason, we also incorporated information on typical $CO/PM_{2.5}$, $NO_x/PM_{2.5}$ and $SO_2/PM_{2.5}$ ratios in the emissions from these, as well as other sources (Marmur et al., 2005). Higher bounds for trace metals are set for gasoline vehicles, compared to diesel vehicles (HEI, 2002; Manchester-Neesvig et al., 2003). For vegetative burning, a constraint of OC/EC$\geqslant$3 was used, as this source is characterized by high OC to EC ratios (US-EPA, 2004b). A relatively large fraction (0.01–0.1) of potassium in biomass burning emissions (US-EPA, 2004b) is also used. For all sources,

the sum of fractions over all species was constrained to be less than or equal to unity. In the case of soil dust and power-plants, oxidized forms of the metals are assumed (such as Al_2O_3, SiO_2 etc.). Organic material (OM) fractions in the primary emissions were bounded by a minimum contribution of 1.4 times the fraction of OC in the profile. These constraints are summarized by the following equations:

$$f_{i,j\ \text{lower}} \leqslant f_{i,j} \leqslant f_{i,j\ \text{upper}}, \tag{2}$$

$$f_{\text{OC},j}/f_{\text{EC},j} \geqslant R_{\text{OC/EC}}, \tag{3}$$

$$f_{\text{OC},j} + f_{\text{EC},j} \geqslant R_{\text{TC}}, \tag{4}$$

$$1.4f_{\text{OC},j} + f_{i(\text{excluding OC}),j} \leqslant 1.0$$
(for all sources but soil-dust and power-plants), (5)

$$1.89f_{\text{Al},j} + 1.40f_{\text{Ca},j} + 1.43f_{\text{Fe},j} + 1.20f_{\text{K},j} + 2.14f_{\text{Si},j} + 1.67f_{\text{Ti},j} + 1.4f_{\text{OC},j} + f_{\text{other},j} \leqslant 1.0$$
(for soil-dust and power-plants), (6)

where, $f_{i,j\ \text{lower}}, f_{i,j\ \text{upper}}$ is the lower and upper bound on fraction of species i in source j (Table 2), $R_{\text{OC/EC}}$ the bound on OC/EC ratio ($\geqslant 1$ for gasoline vehicles; $\geqslant 3$ for vegetative burning; $\leqslant 1$ for diesel vehicles; Table 2), R_{TC} the bound on TC (EC + OC) fraction ($\geqslant 0.5$ for gasoline and diesel vehicles, vegetative burning; Table 2), 1.89, 1.40, 1.43, 1.20, 2.14, 1.67 the ratios of molecular weights of metal-oxide/metal for Al_2O_3, CaO, Fe_2O_3, K_2O, SiO_2 and TiO_2, respectively.

To address the formation of secondary pollutants, four pure component profiles were used for ammonium-sulfate (AMSULF; 73% SO_4^{-2}, 27% NH_4^+), ammonium-bisulfate (AMBSLF; 84% SO_4^{-2}, 16% NH_4^+), ammonium-nitrate (AMNITR; 78% NO_3^-, 22% NH_4^+) and other/secondary OC (OTHROC; 100% OC), based on the molecular weights of the components (Marmur et al., 2005).

For each sample, Eq. (1) was solved by minimizing χ^2:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(C_i - \sum_{j=1}^{n} f_{ij} S_j\right)^2}{\sigma_{C_i}^2}, \tag{7}$$

where σ_{C_i} is the uncertainty of the C_i measurement. This is solved subject to the constraints on the $PM_{2.5}$ source compositions (Eqs. (2)–(6) and Table 2), as well as the requirement to reasonably reconstruct ambient gas-phase (SO_2, CO, and NO_y) concentrations:

$$\frac{1}{b}[\text{GS}] \leqslant \sum_{j=1}^{n} \left(\frac{\text{GS}}{\text{PM}_{2.5}}\right)_j S_j \leqslant b[\text{GS}], \tag{8}$$

where GS is the ambient concentration of gaseous-species (CO, SO_2, NO_y; $\mu g\,m^{-3}$), $\text{GS}\big/(\text{PM}_{2.5})_j$ the mass ratio in emissions from source j (see Marmur et al., 2005 for values used), S_j the contribution (source-strength) of source j ($\mu g\,m^{-3}$) to the $PM_{2.5}$ loading, n the total number of sources, b the bound for gas-species mass reconstruction (typically $b = 3$, to account for uncertainties in initial $GS/PM_{2.5}$ ratios and changes in these ratios during transport from source to receptor; (Marmur et al., 2005)).

This latter requirement has been shown to reduce collinearity between source-compositions and to achieve more plausible source-apportionment results (Marmur et al., 2005).

A global optimization program, Lipschitz global optimizer (LGO) (Pinter, 1996; Pintér, 1997), was utilized to find the optimal solution (by minimizing χ^2), subject to the above mentioned constrains. In LGO, the best solution is sought that satisfies all stated feasibility constraints and maximizes (or minimizes) the value of a given objective function (Pinter, 1996, 1997). The objective of global optimization is to find the best solution of nonlinear decision models, in the possible presence of multiple locally optimal solutions. LGO integrates a suite of robust and efficient global and local scope solvers. These include: global adaptive partition and search (branch-and-bound); adaptive global random search; local (convex) unconstrained optimization; and local (convex) constrained optimization. The LGO implementation of these methods does not require derivative information. Their operations are based exclusively on the computation of the objective and constraint function values, at algorithmically selected search points.

2.2. SEARCH 25 month dataset, Jefferson St., Atlanta, Georgia

Evaluation of this expanded CMB approach involved using the SEARCH (Southeastern Aerosol Research and Characterization) 25 month (8/98–8/00) dataset for the Jefferson St. (JST) monitoring site in Atlanta, GA (Hansen et al., 2006, 2003; Kim et al., 2003; Marmur et al., 2005), which includes data on total $PM_{2.5}$ mass (gravimetric measure) and

its components (major ions by ion chromatography; trace metals by X-ray fluorescence; organic and elemental carbon by thermal optical reflectance), as well as ambient concentrations of SO_2, CO and NO_y. Summation of the analytical uncertainty and $\frac{1}{3}$ of the detection limit value was used as the overall uncertainty assigned to each measured value. Values below the detection limit were replaced by half of the detection limit values, and their overall uncertainties were set at $\frac{5}{6}$ of the detection limit values. Missing values were replaced by the geometric mean of the measured values, and their accompanying uncertainties were set at 4 times this geometric mean value (Marmur et al., 2005). Five variable source-profiles (GV, DV, DUST, BURN, CFPP; Table 2) and four constant ones (AMSULF, AMBSLF, AMNITR, OTHROC) were included in the analysis.

3. Results

3.1. Derived source-profiles

First, source profile compositions were determined (for five sources: GV, DV, DUST, BURN, CFPP; for the species in Table 2) using, initially, a subset of the data. In this way, a separate set of data could be used in the evaluation. Out of the total of 762 days, we identified 447 days in which all of the data (all ions, carbon fractions, metals, CO, SO_2, NO_y) were available. From those 447 days, we chose all the January, April, July and October samples (133 days) to represent the four seasons. LGO then was applied to find the optimal solution Eq. (1) based on the ordinary weighted least-squares (OWLS) approach Eq. (7) (Christensen and Gunst, 2004; Friedlan, 1973) while adjusting the source profile fractions within the allowable bounds (Table 2, Eqs. (2)–(6)) and subject to the gas-phase constraints Eq. (8). This becomes an optimization problem with more than one hundred decision variables (fractions and contributions), requiring several minutes of computational workload (on a Pentium 4.0 PC) and several tens of thousands of iterations per sample to reach a global minimum point. The computational workload for a solution using predetermined source-profiles is much smaller, reaching convergence within several seconds and several hundreds of iterations per sample. Source profiles obtained are analyzed for how often bounds (lower or upper) on individual species are met and for compositional variability between samples. The process was evaluated by repeating the analysis using all the February, May, August (excluding 1998), and November samples (149 days). No significant seasonal variability in source composition was observed. Therefore, the comparison will focus on average source compositions for each of the two test cases.

Average source-profile compositions for the two scenarios (two subsets of the data: Case 1 based on 133 samples; Case 2 based on 149 samples) show little difference (Figs. 1–5). When compared to several source profiles from the literature (Chow et al., 2004; Cooper, 1981; Zielinska et al., 1998) differences arise, but the species driving source apportionment modeling (Marmur et al., 2006) are the same. Major differences are observed for primary sulfate, nitrate and ammonium content in various sources because LGO assigns most of that

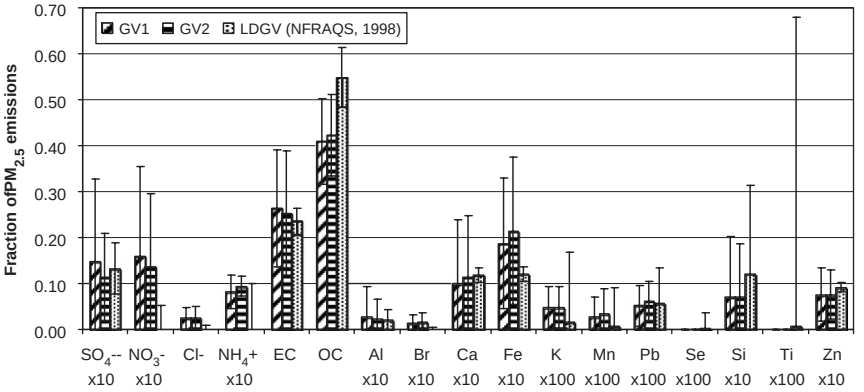

Fig. 1. Source profiles generated by LGO for gasoline-fueled vehicles (GV1 based on 133 cases, GV2 based on 149 cases), compared to a profile from the NFRAQS study (Zielinska et al., 1998), previously used to apportion $PM_{2.5}$ in Atlanta (Marmur et al., 2005). Bars represent $\pm$one standard-deviation of the LGO estimated (over 133 and 149 cases) or measured fractions.

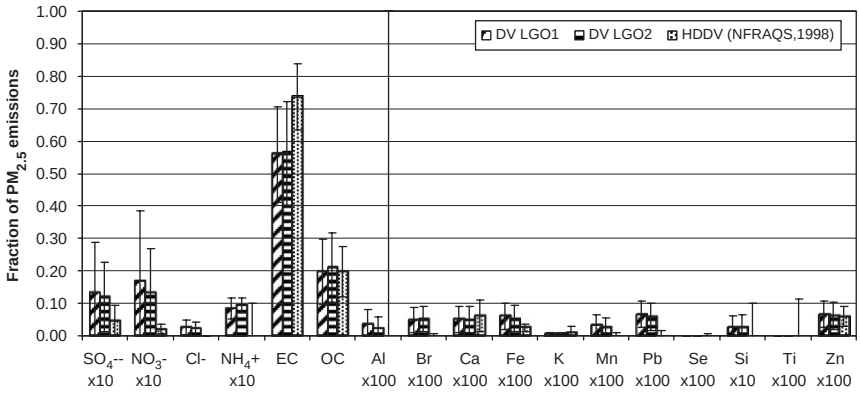

Fig. 2. Source profiles generated by LGO for diesel-fueled vehicles (DV1 based on 133 cases, DV2 based on 149 cases), compared to a profile from the NFRAQS study (Zielinska et al., 1998), previously used to apportion $PM_{2.5}$ in Atlanta (Marmur et al., 2005). Bars represent ± one standard-deviation of the LGO estimated (over 133 and 149 cases) or measured fractions.

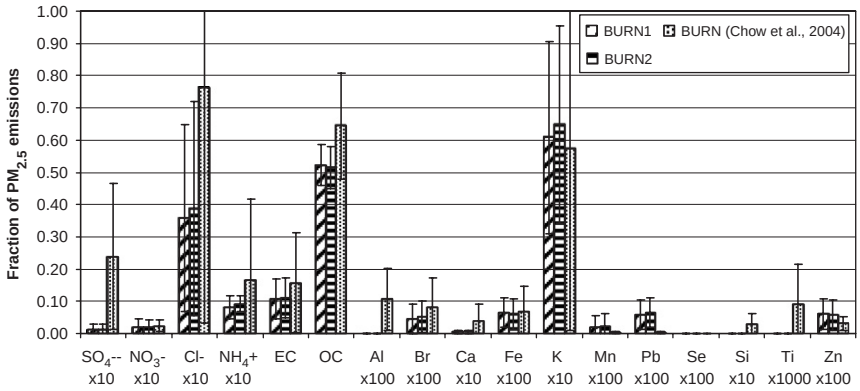

Fig. 3. Source profiles generated by LGO for vegetative burning (BURN1 based on 133 cases, BURN2 based on 149 cases), compared to a vegetative burning profile from the BRAVO study (Chow et al., 2004), previously used to apportion $PM_{2.5}$ in Atlanta (Marmur et al., 2005). Bars represent ± one standard-deviation of the LGO estimated (over 133 and 149 cases) or measured fractions.

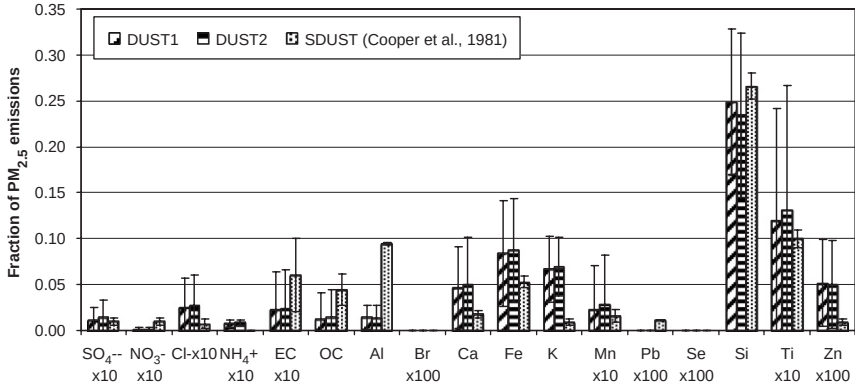

Fig. 4. Source profiles generated by LGO for soil-dust (DUST1 based on 133 cases, DUST2 based on 149 cases), compared to an Alabama soil dust profile (Cooper, 1981), previously used to apportion $PM_{2.5}$ in Atlanta (Marmur et al., 2005). Bars represent ± one standard-deviation of the LGO estimated (over 133 and 149 cases) or measured fractions.

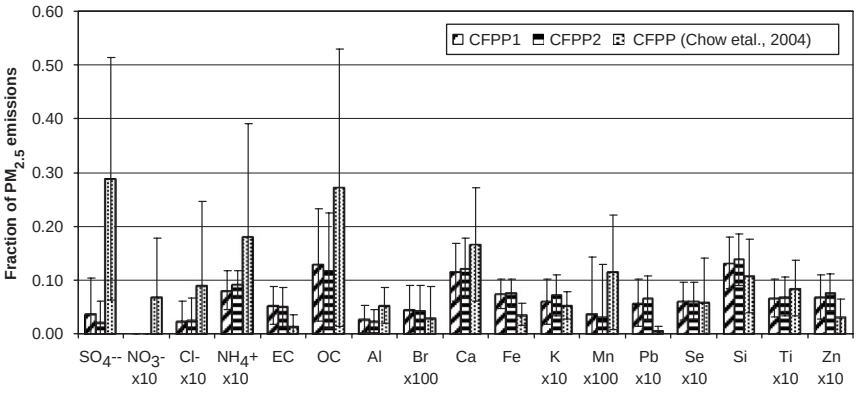

Fig. 5. Source profiles generated by LGO for coal-fired power plants (CFPP1 based on 133 cases, CFPP2 based on 149 cases), compared to a CFPP profile measured in Texas (Chow et al., 2004), previously used to apportion $PM_{2.5}$ in Atlanta (Marmur et al., 2005). Bars represent ± one standard-deviation of the LGO estimated (over 133 and 149 cases) or measured fractions.

mass into the secondary sulfate and nitrate categories. These estimates of primary sulfate and nitrate content are therefore highly uncertain; however, these species do not serve as markers for any of the sources of primary emissions. Hence, the effect of this uncertainty on the overall source-apportionment process is likely a bias in predicted impacts rather than uncertainty in the prediction of day-to-day variation in source impacts.

LGO generated gasoline-vehicle profiles (GV1 and GV2 in Fig. 1) are comparable to the profile from NFRAQS (Zielinska et al., 1998), being characterized by an OC/EC ratio of 1.6 on average (a constraint of $\geqslant 1.0$ was used), compared to 2.3 in the NFRAQS profile. The total carbon content (TC) is 0.67 (a constraint of $\geqslant 0.5$ was used), compared to 0.78, in the NFRAQS profile. Content of Zn, a good marker for gasoline vehicles in the Atlanta source mixture (Marmur et al., 2006; Marmur et al., 2005), is similar (average of 0.008 vs. 0.009). The content of other trace metals (Al, Ca, Fe, Si) is similar.

Diesel vehicle profiles generated by LGO (DV1 and DV2) are characterized by an OC/EC ratio of 0.36 (a constraint of $\leqslant 1.0$ was used) compared to 0.27 in the NFRAQS profile (Zielinska et al., 1998). The TC content is 0.77 on average (a constraint of $\geqslant 0.5$ was used), compared to 0.93. Metal content is similar among all diesel vehicle profiles.

Vegetative burning profiles from LGO were fairly similar to the BURN profile from BRAVO (Chow et al., 2004), with an OC/EC ratio of 4.7 on average, compared to 4.1 from BRAVO. Potassium content in the LGO profiles is 0.063 on average, compared to 0.056 in the BRAVO profile. The LGO profiles' chlorine content is roughly half of that in the measured profile (0.037 compared to 0.076).

The LGO-derived soil dust profiles are similar to the Alabama soil-dust profile from Cooper et al. (Cooper, 1981) with respect to Si, Ti, Mn and Fe content, but Al, Ca and K content differed significantly. The high Al content in the Alabama profile (Cooper, 1981) seems to be an overestimate for Atlanta aerosol (Marmur et al., 2005).

Some differences arise when comparing the LGO derived CFPP profile to the one from BRAVO (Chow et al., 2004), which are based on measurements in Texas. SO_4^{-2}, EC, OC, Al, Ca, Fe and Si are the most abundant species in both sets of profiles, but differences in their content is evident, especially in SO_4^{-2} and OC content, though neither is an important tracer for primary CFPP $PM_{2.5}$. As previously mentioned, LGO assigns most of the SO_4^{-2} to the secondary "ammonium-sulfate" category, likely underestimating sulfate content in primary emissions. OC is apportioned to carbon-rich source-categories such as GV, DV and BURN based on constraints on both OC content and OC/EC ratios, and to the "other OC" category (secondary and un-apportioned organic carbon). For relatively carbon-lean source-categories, such as DUST and CFPP, for which knowledge on typical OC/EC ratios is limited, LGO tends to suggest a lower fraction of OC. The content of selenium, a unique tracer for CFPP, is very similar, 0.0061 in the LGO profiles, compared to 0.0058 in the BRAVO profile.

The role of the constraints/bounds used to derive the source compositions were analyzed using the

Table 3
Percent of cases (447 total) in which the derived species fraction was at the lower limit/within the allowable range/at the upper limit

Species	LDGV (low/within/upper)	HDDV (low/within/upper)	BURN (low/within/upper)	DUST (low/within/upper)	CFPP (low/within/upper)
EC	20/80/0	23/77/0	18/82/0	63/17/20	25/52/23
OC	0/100/0	35/65/0	0/100/0	66/27/7	19/79/3
OC/EC	15/85/0	0/100/0	3/97/0	—	—
TC	1/99/0	4/96/0	1/99/0	—	—
SO_4	92/7/1	90/8/2	89/9/2	84/14/2	77/22/1
NO_3	88/9/3	82/13/4	83/10/7	78/15/7	—
Cl	45/52/2	39/59/2	41/50/9	66/20/14	59/25/16
NH_4	11/7/82	11/6/83	11/11/78	13/18/69	11/13/76
Al	64/35/1	51/31/19	—	80/20/0	53/45/3
As	—	—	—	—	1/11/88
Ba	—	—	—	7/15/78	2/12/87
Br	18/81/1	30/48/23	34/34/32	—	37/33/29
Ca	30/66/3	36/39/25	38/27/35	33/64/3	3/87/10
Fe	14/78/8	32/40/28	26/32/42	14/81/6	6/63/32
K	35/33/32	38/23/39	6/75/19	14/57/28	30/25/46
Mn	54/45/0	46/48/6	55/35/9	66/34/0	66/31/3
Pb	20/39/41	15/48/37	23/28/49	—	19/34/47
Se	—	—	—	—	2/71/26
Si	44/53/3	41/49/10	—	11/89/0	8/81/11
Ti	—	—	—	23/74/3	15/41/44
Zn	4/89/7	22/42/36	25/30/45	32/27/41	19/25/56

entire dataset of derived source-profiles (447 cases), in terms of percent of cases in which either bound (lower/upper) were found limiting (Table 3). These data indicate that the constraints are most often limiting for species that are not unique tracers or key driving species of a given category. Examples are sulfate, nitrate and ammonium content in all sources of primary $PM_{2.5}$, OC and non-crustal elements in soil dust, and various metals such as Al, K, Mn and Si in both types of mobile sources. However, a key success of the source-profile derivation process is LGO's ability to estimate the fraction of unique/key species well within the allowable range in most cases. Examples are EC, OC, TC, OC/EC ratio and Zn for gasoline vehicles; EC, OC, TC and OC/EC ratio for diesel vehicles; EC, OC, TC, OC/EC ratio and K for wood burning; Ca, Fe, Si, and Ti for dust; and Ca and Se for coal-fired power plants.

3.2. Source apportionment based on the derived $PM_{2.5}$ source profiles

Using LGO-derived source profiles (LDSP) based on all available samples (447 cases for the period of 1/8/1998–31/8/2000, Table 4) to apportion daily $PM_{2.5}$ levels measured at the Jefferson Street site in Atlanta, typically led to similar results as when measurement-based source profiles (MBSP) were used (Marmur et al., 2005) (Fig. 6; Tables 5–7), though with a couple major differences.

$PM_{2.5}$ attributed to wood burning was $0.66\,\mu g\,m^{-3}$, on average, using LDSP versus $1.1\,\mu g\,m^{-3}$ using MBSP. This is driven, in part, by the higher potassium fraction in the LGO derived DUST profile, compared to the measurement-based DUST profile. Other differences include diesel $PM_{2.5}$ ($2.3\,\mu g\,m^{-3}$ using LDSP, $1.9\,\mu g\,m^{-3}$ using MBSP) and "other OC" (3.1 and $2.5\,\mu g\,m^{-3}$, respectively).

Comparing the quality of fit achieved in the two cases (Table 5), finds a significantly lower χ^2 value (error function being minimized) (Marmur et al., 2005) using LDSP (12.6) compared to MBSP (20.3). This is driven by several trace species, such as Br, Ca, Fe, K, Pb and Si, for which their ambient concentrations were better reconstructed using LDSP (Al was not used as a fitting species in the MBSP solution). However, EC, Cl^- and Zn were better fit using the MBSP. The improved fit for potassium using LDSP can partially explain the lower mass contribution of BURN using LDSP, compared to MBSP. The improved fit for Si, Fe and Al using LDSP may indicate that the DUST profile

Table 4
LGO derived $PM_{2.5}$ source profiles based on 447 days in which all relevant data (ions, EC, OC, metals, CO, SO_2, NO_y) were available

Species	GV	DV	DUST	BURN	CFPP
SO_4^{-2}	0.0129±0.0138	0.0128±0.0139	0.0013±0.0015	0.0013±0.0016	0.0307±0.0563
NO_3^-	0.0144±0.0174	0.0161±0.0205	0.0001±0.0003	0.0018±0.0025	0.0000±0.0000
Cl^-	0.0240±0.0236	0.0238±0.0222	0.0026±0.0033	0.0374±0.0320	0.0023±0.0039
NH_4	0.0088±0.0031	0.0088±0.0030	0.0008±0.0004	0.0086±0.0032	0.0087±0.0031
EC	0.2575±0.1323	0.5654±0.1570	0.0024±0.0041	0.1093±0.0609	0.0522±0.0357
OC	0.4176±0.0914	0.2063±0.1059	0.0150±0.0301	0.5225±0.0626	0.1280±0.1036
Al	0.0032±0.0073	0.0003±0.0004	0.0150±0.0154	0.0000±0.0000	0.0253±0.0238
As	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0010±0.0002
Ba	0.0000±0.0000	0.0000±0.0000	0.0089±0.0028	0.0000±0.0000	0.0095±0.0019
Br	0.0014±0.0018	0.0005±0.0004	0.0000±0.0000	0.0005±0.0005	0.0004±0.0005
Ca	0.0109±0.0133	0.0005±0.0004	0.0467±0.0469	0.0005±0.0005	0.1157±0.0552
Fe	0.0210±0.0157	0.0005±0.0004	0.0867±0.0582	0.0006±0.0005	0.0745±0.0280
K	0.0005±0.0005	0.0001±0.0000	0.0668±0.0342	0.0628±0.0304	0.0066±0.0041
Mn	0.0003±0.0007	0.0003±0.0003	0.0025±0.0046	0.0002±0.0004	0.0004±0.0011
Pb	0.0006±0.0004	0.0006±0.0004	0.0000±0.0000	0.0006±0.0005	0.0063±0.0043
Se	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0061±0.0035
Si	0.0076±0.0130	0.0026±0.0037	0.2419±0.0897	0.0000±0.0000	0.1341±0.0475
Ti	0.0000±0.0000	0.0000±0.0000	0.0120±0.0125	0.0000±0.0000	0.0069±0.0036
Zn	0.0074±0.0054	0.0006±0.0004	0.0005±0.0005	0.0006±0.0005	0.0075±0.0037

GV—gasoline vehicles; DV—diesel vehicles; BURN—vegetative burning; DUST—soil dust; CFPP—coal fired power plants.

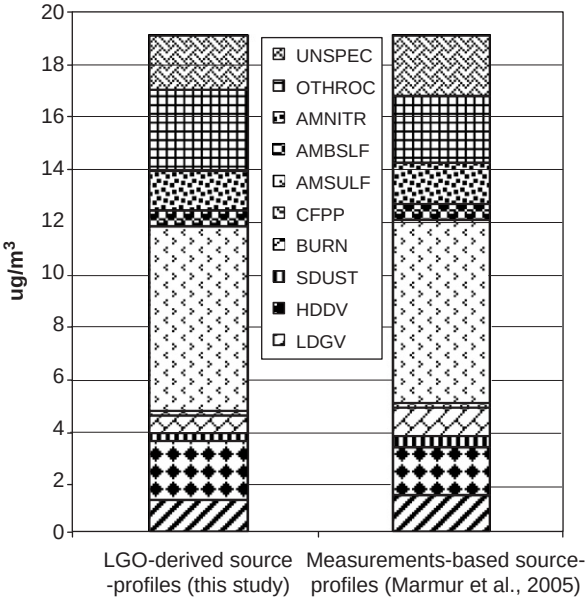

Fig. 6. Average source contributions (1/8/98–31/8/00) to $PM_{2.5}$ at the Atlanta Jefferson Street site, using LGO-derived source-profiles (this study) and measurement-based source-profiles (Marmur et al., 2005) (GV—gasoline vehicles; DV—diesel vehicles; BURN—vegetative burning; DUST—soil dust; CFPP—coal fired power plants; AMSULF—ammonium sulfate; AMBSLF—ammonium-bisulfate; AMNITR—ammonium-nitrate; OTHROC—Other OC; UNSPEC—unspecified).

derived by LGO is more representative of Atlanta soil dust, compared to the Alabama soil profile used (Cooper, 1981). Aluminum was excluded as a fitting species in the MBSP study (Marmur et al., 2005) because including it did not improve the fit significantly (calculated/observed ratio for Al was 4.0 when included, 4.7 when excluded), but the chi-square increased (22.4 vs. 20.3), indicating that the DUST impact was driven by another species (Si) (Marmur et al., 2006), and that the Al/Si ratio in the soil profile is too high compared to ambient measurements in Atlanta.

To assess the difference in daily variability in source impacts based on LDSP and MBSP, we also computed correlations between the various source-contribution estimates (Table 6). Of the five source categories for which profiles have been derived, the source inter-correlations are high for DUST (0.97), GV (0.93), and CFPP (0.89), slightly lower for BURN (0.83), and relatively low for DV (0.68). DV and BURN were previously mentioned for differences in their average source contributions based on the two methods (Fig. 6). The correlations for all the secondary $PM_{2.5}$ categories are high (0.95–1.00).

Further analyzing differences in the DV and BURN source contributions predicted by the two methods, MBSP DV source impact is highly correlated with EC (0.96) whereas the LDSP DV source impact has a lower correlation with EC (0.72). This indicates that EC is more of a driving force in the MBSP solution than in the LDSP solution. For BURN, the correlations with K were

Table 5
Performance measures for the LDSP and MBSP solutions

	LDSP (this study)	MBSP (Marmur et al., 2005)
	Mean (S.D.)	Mean (S.D.)
Chi-square (error function)	**12.4** (12.0)	20.3 (16.8)
R	0.9836 (0.0349)	0.9879 (0.0324)
% Total mass	92.3 (18.7)	90.5 (17.4)
SO_4^{-2} ratio	1.06 (0.07)	1.07 (0.07)
NO_3^- ratio	1.16 (0.79)	1.18 (0.87)
Cl^- ratio	1.31 (0.77)	**1.06** (0.63)
NH_4^+ ratio	0.88 (0.15)	0.88 (0.15)
EC ratio	0.94 (0.37)	**0.98** (0.13)
OC ratio	1.00 (0.03)	1.00 (0.03)
Al ratio[a]	**1.50** (0.86)	4.67 (2.81)
As ratio	0.19 (0.18)	0.32 (0.25)
Ba ratio	0.26 (0.20)	0.10 (0.08)
Br ratio	1.26 (1.36)	0.39 (0.38)
Ca ratio	1.08 (0.34)	1.15 (0.34)
Fe ratio	**0.85** (0.19)	0.55 (0.17)
K ratio	**1.05** (0.46)	1.19 (0.49)
Mn ratio	1.57 (1.23)	0.69 (0.58)
Pb ratio	**1.12** (0.95)	0.27 (0.23)
Se ratio	1.25 (1.27)	**1.11** (1.20)
Si ratio	**1.07** (0.19)	1.28 (0.12)
Ti ratio	1.16 (0.70)	1.27 (0.70)
Zn ratio	0.92 (0.37)	**1.01** (0.35)
SO_2 ratio	2.13 (0.89)	1.99 (0.97)
CO ratio	1.75 (0.84)	2.06 (0.83)
NO_y ratio	1.63 (0.68)	1.58 (0.66)

Chi-square (error function), correlation (*R*) between ambient and reconstructed $PM_{2.5}$, percent of total mass explained, and calculated-to-observed ratios (ideally would approach 1). Bolded values are superior compared to the other solution.

[a]Al was not included as a fitting species in the MBSP solution (Marmur et al., 2005). To allow for a full comparison, we also reran the MBSP analysis with Al as fitting species. Al fit was slightly improved (calculated/observed ratio of 4.0 vs. 4.7), though still very much overestimated, while the overall fit (as expressed by chi-square) worsened (22.4 vs. 20.3).

more similar, 0.62 based on the MBSP solution, 0.67 based on LDSP.

Effects of fluctuations in tracer concentrations on source contributions are investigated further by a sensitivity analysis, in which the ambient concentrations of one $PM_{2.5}$ component at a time were increased by 50% and the resulting effects on the source-attributions (using the fixed derived source profiles) were analyzed. These results are compared to a similar analysis performed on the MBSP solution (Marmur et al., 2006) (Table 7). DV source contributions are driven mainly by EC in both solutions (62% and 70% increase in DV contribution based on LDSP and MBSP due to a 50% increase in EC concentrations), but the LDSP is less sensitive to EC and more sensitive to Si compared to the MBSP solution. In addition, EC has a bigger effect on the spilt between gasoline and diesel vehicles in the LDSP solution, scavenging more mass from the GV category compared to the MBSP case. Similarly, mass is scavenged from the BURN category using the LDSP, not so using MBSP. The BURN impact based on the LDSP solution is more sensitive to K.

The selection of bounds is a critical step in the analysis, and this choice can have an effect on the solutions obtained. However, setting the bounds based on well-based knowledge of typical source compositions (such as in Tables 1, 2) reduces the possibility of noise or randomness in the source-attributions. To assess the effect of bound selection on the source-apportionment results, we repeated the analysis, this time relaxing both the lower (dividing) and upper (multiplying) species fraction bounds by a factor of two. This had little effect on the temporal patterns in source contributions, with

Table 6
Correlation matrix (*R*) of source-contributions based on LDSP and MBSP solutions

		LDSP (this study)								
		GV	DV	DUST	BURN	CFPP	AMSULF	AMBSLF	AMNITR	OTHROC
MBSP (Marmur et al., 2005)	GV	**0.93**	0.43	0.12	0.33	0.12	0.07	−0.02	0.30	0.29
	DV	0.22	**0.68**	0.19	0.62	0.22	0.26	−0.03	0.21	0.41
	DUST	0.17	0.33	**0.97**	0.17	0.22	0.33	−0.08	−0.08	0.17
	BURN	0.10	0.02	−0.15	**0.83**	−0.08	−0.04	0.05	0.22	0.12
	CFPP	0.28	0.39	0.15	0.16	**0.89**	0.21	0.06	0.08	0.31
	AMSULF	0.03	0.23	0.29	0.13	0.24	**1.00**	−0.08	−0.03	0.30
	AMBSLF	0.03	−0.04	−0.07	0.05	0.07	−0.12	**0.95**	0.17	0.01
	AMNITR	0.26	0.15	−0.09	0.25	0.02	−0.05	0.18	**0.98**	0.11
	OTHROC	0.42	0.70	0.20	0.28	0.30	0.29	0.03	0.14	**0.97**

Bolded values represent same-source correlations.

Table 7
Percent change in average source-attributions for a 50% increase in concentrations of several tracer species (increased one at a time)

Species	% change in source-attribution for a 50% increase in the corresponding species-concentration (LDSP/MBSP)					
	GV	DV	DUST	BURN	CFPP	OTHROC
EC	−17/−6.8	62/70	2.8/1.2	−6.7/0	0.5/−1.2	−5.1/−5.6
OC	0/0.6	0/0	0/0	−0.1/0.6	0/0	73/82
Al	0.6/0	−0.1/0.3	0.5/0.7	−0.4/1	0.8/1.4	−0.1/1.7
Br	4.8/0	3.1/−0.5	−1.2/−0.5	4.1/2.8	−2.7/0	−1.5/−0.9
Ca	1.5/1.3	−0.5/−1.3	1.7/−2	−2.7/1.3	8.1/29	−0.1/−0.7
Fe	6.7/0	−4.9/0	14.3/0	−13.9/0	−0.9/0	0.1/0
K	−4.7/0	−5.6/−8.5	1.3/−1.7	110/40	0.7/−3.4	−7.9/−12
Mn	0.4/0	0.6/0	4.2/3	−1.1/0	−1.4/0	−0.2/0
Pb	0/1.0	0.1/−0.3	−0.7/0	1.2/0.3	0.7/−0.3	−0.3/−0.3
Se	−1.5/−0.9	0.5/0	−1.2/−1.4	0.8/0	3.1/8.6	0.2/0
Si	−5.0/−1.0	5.3/0.7	27/37	−21/−1.4	−1.8/−5.5	1.5/0.3
Ti	−1.5/−0.9	0.7/0	6.6/7.3	−2.4/0	0/0.9	0.6/0
Zn	38/13	−6.9/−4	−6.1/−2.3	6/0.6	−6.2/−2.8	−2.9/−2.8
SO_2	−2.5/−0.8	0.4/0.5	−7/−4.6	4.8/1.3	48/32	−0.2/−0.3
CO	7.0/5.6	−1.8/−2.1	0/−0.7	0/0	−1.5/−2.8	−0.3/−1.4
NO_y	3.2/3.4	0.9/1.7	−2.6/−0.6	−0.3/−0.6	−2.9/−1.1	−0.4/−0.6

Table 8
Comparison between source apportionment results (total mass and correlations) based on PMF (Kim et al., 2004a), LSDP (this study), and MBSP (Marmur et al., 2005)

Source category	LDSP/PMF mass ratio	MBSP/PMF mass ratio	Correlation (R): PMF, LDSP	Correlation (R): PMF, MBSP
Gasoline	1.15	1.31	0.52	0.50
Diesel	1.23	0.98	0.72	0.78
Soil	0.57	0.72	0.93	0.97
Wood	0.62	1.00	0.71	0.78

source inter-correlations of 0.93, 0.88, 0.99, 0.90, and 0.97 for GV, DV, DUST, BURN, and CFPP, respectively, for the sensitivity and baseline cases. Inter-correlations for the four secondary $PM_{2.5}$ categories were near perfect (0.98–1.00). Average mass attributions changed by 10% or less for all source categories except GV (23% reduction compared to base-case) and DV (31% reduction). To assess whether solutions obtained are unique, we also repeated the analysis this time changing the optimization starting point. Changing the initial estimate of the mass apportioned to each category (while keeping source compositions fixed) had no effect on the final results obtained, but changing the initial estimate of the source compositions (from the midpoint of the allowed range to the extreme) did change the results slightly, with source inter-correlations of 0.95 and above for all sources except for GV (0.88).

3.3. Comparison with FA results

Both sets of source apportionment results (LDSP and MBSP) provide reasonable estimates of the impacts of various sources on ambient $PM_{2.5}$ levels, as reflected by the sensitivity analysis results and correlations with ambient tracer concentrations (Marmur et al., 2005, 2006). While there is no standard by which to compare the accuracies of the LDSP and MBSP results, a comparison of these results with FA results provides an indication of consistency across methods. We compare the LDSP and MBSP results to those from a PMF study (Kim et al., 2004a) for the JST site for the period of 11/

98–8/00 (Table 8). Only overlapping source categories are compared (gasoline, diesel, soil, wood). There is more agreement between the results based on PMF and MBSP than PMF and LDSP for diesel vehicles, wood burning and soil dust. This is expressed by both the average mass apportioned to the various categories and by the correlations between the various source-apportionment methods (Table 8). Impacts of gasoline vehicles and the split between diesel and gasoline vehicles (diesel/gasoline ratio of 1.6, 1.7 and 1.2 based on PMF, LDSP and MBSP results, respectively) are more alike in PMF and LDSP results. The overall greater agreement between PMF and the MBSP results is surprising given that in theory LDSP and PMF are more alike. However, the differences between method intercorrelations (Table 8) are minor, and are a reflection of the small differences between the LDSP and MBSP results. These results might also suggest that the differences between CMB and FA results are more due to what the sources and factors, respectively, represent than due to the accuracy of the source profiles used in CMB.

4. Conclusions

Results from an expanded CMB approach deriving source-compositions based on ambient data were compared with CMB results based on measured source-profiles. For most sources, there is substantial agreement between the two methods. Despite overall lower residual mass obtained by the expanded approach presented here, there is no standard by which to compare the accuracy of these two methods, especially in how well they capture the temporal trends in source impacts. As such, the approach presented here can be viewed as one method to assess the representativeness of measured source-profiles and to help identify those profiles that may be in significant error. It can also be used to quantify uncertainties in source-impact estimates, which are in part due to uncertainties in source compositions.

Acknowledgements

This work was supported by Grants to Emory University from the US Environmental Protection Agency (R82921301-0) and the National Institute of Environmental Health Sciences (R01ES11199 and R01ES11294) and to Georgia Tech (EPA Grants R831076 and R830960). We would also like to thank ARA (Atmospheric Research and Analysis) for providing access to data.

References

Chow, J.C., Watson, J.G., Kuhns, H., Etyemezian, V., Lowenthal, D.H., Crow, D., Kohl, S.D., Engelbrecht, J.P., Green, M.C., 2004. Source profiles for industrial, mobile, and area sources in the big bend regional aerosol visibility and observational study. Chemosphere 54 (2), 185–208.

Christensen, W.F., Gunst, R.F., 2004. Measurement error models in chemical mass balance analysis of air quality data. Atmospheric Environment 38 (5), 733–744.

Cooper, J.A., 1981. Determination of source contributions to fine and coarse suspended particulate levels in Petersville, Alabama. Report to Tennessee Valley Authority, NEA Inc.

Friedlan, S., 1973. Chemical element balances and identification of air-pollution sources. Environmental Science and Technology 7 (3), 235–240.

Gillies, J.A., Gertler, A.W., 2000. Comparison and evaluation of chemically speciated mobile source PM2.5 particulate matter profiles. Journal of The Air and Waste Management Association 50 (8), 1459–1480.

Hansen D.A., Edgerton E.S., Hartsell B.E., Jansen J.J., Hidy G.M., 2006. Air quality measurements for the aerosol research and inhalation epidemiology study. Journal of the Air and Waste Management Association, in press.

Hansen, D.A., Edgerton, E.S., Hartsell, B.E., Jansen, J.J., Kandasamy, N., Hidy, G.M., 2003. The southeastern aerosol research and characterization study: part 1—overview. Journal of the Air and Waste Management Association 53 (12), 1460–1471.

HEI, 2002. Emissions from Diesel and Gasoline Engines Measured in Highway Tunnels, vol. 107.

Hopke, P.K., 1988. Target transformation factor-analysis as an aerosol mass apportionment method—a review and sensitivity study. Atmospheric Environment 22 (9), 1777–1792.

Kim, E., Hopke, P.K., Edgerton, E.S., 2003. Source identification of Atlanta aerosol by positive matrix factorization. Journal of The Air and Waste Management Association 53 (6), 731–739.

Kim, E., Hopke, P.K., Edgerton, E.S., 2004a. Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization. Atmospheric Environment 38 (20), 3349–3362.

Kim, E., Hopke, P.K., Larson, T.V., Maykut, N.N., Lewtas, J., 2004b. Factor analysis of Seattle fine particles. Aerosol Science And Technology 38 (7), 724–738.

Manchester-Neesvig, J.B., Schauer, J.J., Cass, G.R., 2003. The distribution of particle-phase organic compounds in the atmosphere and their use for source apportionment during the southern California children's health study. Journal of The Air and Waste Management Association 53 (9), 1065–1079.

Marmur, A., Park, S.K., Mulholland, J.A., Tolbert, P.E., Russell, A.G., 2006. Source apportionment of PM2.5 in the southeastern United States using receptor and emissions-based models: conceptual differences and implications for time-series health studies. Atmospheric Environment 40, 2533–2551.

Marmur, A., Unal, A., Mulholland, J.A., Russell, A.G., 2005. Optimization based source apportionment of PM2.5 incorporating gas-to-particle ratios. Environmental Science and Technology 39 (9), 3245–3254.

Maykut, N.N., Lewtas, J., Kim, E., Larson, T.V., 2003. Source apportionment of PM2.5 at an urban IMPROVE site in Seattle, Washington. Environmental Science and Technology 37 (22), 5135–5142.

Paatero, P., 2004. User's Guide for Positive Matrix Factorization Programs PMF2 and PMF3, part 1: tutorials.

Paatero, P., Tapper, U., 1994. Positive matrix factorization—a nonnegative factor model with optimal utilization of error-estimates of data values. Environmetrics 5 (2), 111–126.

Pinter, J.D., 1996. Global Optimization in Action. Kluwer Academic Publishers, The Netherlands.

Pintér, J.D., 1997. LGO—a program system for continuous and Lipschitz global optimization. In: Developments in Global Optimization. Kluwer Academic Publishers, Dordrecht, Boston, London.

Seinfeld, J.H., Pandis, S.N., 1998. Atmospheric Chemistry and Physics. Wiley, New York.

US-EPA, 2004a. EPA-CMB8.2 Users Manual (EPA-452/R-04-011), Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711. Available at 〈http://www.epa.gov/scram001/models/receptor/〉.

US-EPA, 2004b. Protocol for Applying and Validating the CMB Model for PM2.5 and VOC (EPA-451/R-04-001), Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711.

Zielinska, B., McDonald, J.D., Hayes, T., Chow, J.C., Fujita, E.M., Watson, J.G., 1998. Northern Front Range Air Quality Study Final Report, vol. B. Source Measurements. Available at 〈http://www.nfraqs.colostate.edu/nfraqs/index2.html〉.